

# TECNOLOGIAS LOD E A PUBLICAÇÃO E INTERLIGAÇÃO DE ACERVOS DIGITAIS DE ARQUIVOS, BIBLIOTECAS E MUSEUS NA WEB

**Carlos Henrique Marcondes** (UFF) - marcon@vm.uff.br

## **Resumo:**

*A partir das limitações dos atuais sistemas gerenciadores de catálogos o trabalho discute as potencialidades das tecnologias de dados abertos interligados para a publicação, interligação e integração ao outros recursos Web, de acervos digitais de arquivos, bibliotecas e museus.*

**Palavras-chave:** *catálogos, arquivos, bibliotecas, museus, dados abertos interligados, Web Semântica*

**Eixo temático:** *Eixo 6: IV EEPC - Encontro de Estudos e Pesquisas em Catalogação. Organização e Tratamento da Informação: tecnologias e novas ferramentas, instrumentos, processos, produtos e serviços, políticas, cooperação.*

## **Eixo 6 IV EEPC - Encontro de Estudos e Pesquisas em Catalogação**

**Resumo:** A partir das limitações dos atuais sistemas gerenciadores de catálogos o trabalho discute as potencialidades das tecnologias de dados abertos interligados para a publicação, interligação e integração aos outros recursos Web, de acervos digitais de arquivos, bibliotecas e museus.

**Palavras-chave:** catálogos, arquivos, bibliotecas, museus, dados abertos interligados, Web Semântica

### **1. Introdução**

Arquivos, bibliotecas e museus têm sido guardiões da memória e cultura desde a Antiguidade. Catálogos têm sido os instrumentos através dos quais essas instituições de memória e cultura divulgam seus acervos. Catálogos passaram de listas alfabéticas de títulos, fichários desdobrados por títulos, autores e assuntos até, com o advento da tecnologia dos computadores (a partir da década de 1960), nos catálogos em linha.

Com o surgimento da Web (cerca de 2000) os catálogos passaram a estar disponíveis, podendo ser consultados a partir de qualquer lugar e a qualquer hora. Acervos também vêm sendo digitalizados e disponibilizados em linha. No entanto a Web também trouxe sérios desafios para essas instituições. Enquanto no mundo não digital elas eram as únicas fontes de informação, hoje sofrem a concorrência de outros recursos como Google Acadêmico, Google Books, YouTube, Amazon, Wikipedia, etc.

Esta situação tem colocado em questão o papel histórico dessas instituições. Profissionais de informação/curadores de acervos de instituições de memória e cultura vêm, de forma crescente, percebendo esses desafios e discutindo alternativas para superá-los (DEMPSEY, 2006), (GORMAN, 2003), (HILLMANN, 2008) (MANN, 2006).

A Web vem se desenvolvendo segundo a visão da Web Semântica (BERNERS-LEE, 2001). Para além de uma Web de documentos hipertextuais interligadas entre si, teríamos uma Web de *recursos*, qualquer coisa que esteja disponível na Web, como documentos hipertextuais, imagens estáticas e em movimento, arquivos sonoros, etc., mas também representações digitais de *coisas* reais como eu, você, um modelo de carro, a Mona Lisa de Da Vinci, o Pão de Açúcar, etc.; é a chamada *Web das coisas*.

As tecnologias LOD – “linked open data”, dados abertos interligados - são parte da Web Semântica e uma nova forma de representar conteúdos que permite a programas, mais que simplesmente publicarem estes conteúdos, também “compreenderem seus significados”, podendo processá-los de forma mais “inteligente”.

Se comparada à proposta da Web Semântica a tecnologia atual dos sistemas de catálogos apresenta grandes limitações, ao restringir o acesso e a interligação dos conteúdos mantidos ao escopo do sistema, transformando este em um “silo” que aprisiona a informação. Além disso, os conteúdos dos acervos mantidos no catálogo em linha só fazem sentido, só têm “semântica”, no escopo desse sistema, seu significado e inteligibilidade estão restritos a este escopo. Um usuário tem que acessar o sistema de catálogo, fazer “login” nele para acessar seus registros. É o sistema de catálogo que recupera registros MARC, com seus identificadores de conteúdo, que só fazem sentido para catalogadores que os conheçam, e os exibe na tela em campos que fazem sentido a um usuário, como autor, título, editor, data, etc. Essa tecnologia limita assim as possibilidades de integração de acervos digitais entre si e aos demais

conteúdos da Web, encerrando esses conteúdos em um mundo próprio, fechado, isolado dos outros fluxos de informação da Web.

A tecnologia dos catálogos, baseada em arquivos invertidos e com algoritmos para resolver operadores booleanos, resolve consultas do tipo:

- Dado um assunto ou um autor, recupere referências sobre esse assunto, ou desse autor.

Mas essa tecnologia é incapaz de resolver consultas do tipo:

- Recupere referências sobre um dado assunto, de autores filiados a universidades e cursos de pós-graduação com conceito CAPES 7.

Arquivos, bibliotecas e museus pelo mundo afora veem desenvolvendo projetos de publicação de seus catálogos usando LOD. Entre os mais significativos estão os da Biblioteca Nacional da Espanha (<http://datos.bne.es>), da British Library (<http://bnb.data.bl.uk>), da Deutsche National Bibliothek ([http://www.dnb.de/EN/Service/DigitaleDienste/LinkedData/linkeddata\\_node.html](http://www.dnb.de/EN/Service/DigitaleDienste/LinkedData/linkeddata_node.html)), da Biblioteca Nacional da França (<http://data.bnf.fr>), da Biblioteca Europeia (<http://labs.europeana.eu/api/linked-open-data-introduction>), do British Museum (<https://datahub.io/dataset/british-museum-collection>), do Archives Hub (<https://archiveshub.jisc.ac.uk/>). Mais que simplesmente publicar esses conteúdos, é necessário dar-lhes visibilidade e integrá-los aos fluxos de informação da Web, tanto criando “links” entre acervos complementares como fazendo com que sejam criados “links” para esses acervos desde outras páginas fora do domínio específico das instituições de memória e cultura.

O objetivo do trabalho é apresentar as tecnologias LOD e discutir suas potencialidade e aplicação na publicação e interligação de acervos digitais de arquivos, bibliotecas e museus. Em especial foram examinadas as metodologias de descrição de acervos em arquivos, bibliotecas e museus para responder como, a partir dos acervos digitais, publicar e interligar objetos de *diferentes* acervos/instituições.

## 2. Método da pesquisa:

Para elaborar e apresentar sistematicamente os tópicos propostos foi utilizado a pesquisa documental e bibliográfica. Manuais como Heath e Bizer (2011) e RDF Primer (2004), e documentos em linha e relatórios do W3C sobre o tema foram consultados. A pesquisa usou como fontes o Google Acadêmico e anais de congressos da IFLA e ICOM, e como estratégia de busca “*linked open data*” AND (*archive OR library OR museum*). Metodologias de descrição de acervos em arquivos, bibliotecas e museus como MARC (MARC STANDARDS, 2017), MODS, Dublin Core, RDA (2010), ISAD(G), LIDO, e modelos conceituais como FRBR (IFLA, 1998), CIDOC CRM (2014), Ric-CM (INTERNATIONAL COUNCIL ON ARCHIVES, 2016), EDM (EUROPEANA DATA MODEL DOCUMENTATION), também foram avaliados para identificar relações, implícitas ou explícitas, que pudessem gerar “links” entre objetos dos acervos ou destes com outros recursos Web, a partir dos catálogos.

## 3. Resultados

As tecnologias LOD consistem em: a-publicar recursos diretamente na Web, associando a estes recursos metadados legíveis por programas representados do esquema RDF. A associação a metadados se dá através de proposições na forma de

<sujeito> (o recurso a ser descrito), <predicado> (uma propriedade do recurso que o descreva) e <objeto> (o valor daquela propriedade para o recurso. Por exemplo, a página <<http://www.uff.br>> tem como <autor> <Marcondes>. Um recurso pode ser desde uma página hipertextual HTML como na Web convencional até um arquivo de som mp3, de imagem estática ou de vídeo, até “coisas”, como eu, você, o Pão de Açúcar, um modelo de automóvel, que tenham representações digitais na Web; é a chamada “Web das coisas”. b- identificar esses recursos para torna-los acessíveis através de identificadores permanentes, IRIs, imunes as quebras de “links” tão comuns na Web atual, e c- especificarem a semântica dos conteúdos de forma precisa e inequívoca, usando vocabulários específicos, também disponíveis publicamente na Web.

Seu potencial para instituições de memória e cultura vem sendo crescentemente destacado (W3C Library Linked Data Incubator Group, 2011), (HANNEMANN, 2010), (ACENJO-BULLÓN, 2015).

Como seriam publicados Itens de acervos digitais como *recursos segundo as tecnologias LOD*? Um item de um acervo digital seria uma versão digitalizada de um livro raro, ou uma fotografia de uma peça museológica, agregado a metadados que o descrevam e o contextualizem, publicado e acessado através de um IRI.

#### 4. Discussão

“The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data” (BERNERS-LEE, 2006).

Embora existam várias experiências de publicação de acervos em arquivos, bibliotecas e museus utilizando as tecnologias LOD, estas experiências têm se limitado basicamente a publicar o conteúdo de catálogos; são poucas as que utilizam essas tecnologias para tornar interoperáveis diferentes acervos. As tecnologias LOD mostram significativas vantagens em termos de interoperabilidade (MARCONDES, 2016).

Metodologias de descrição de acervos em arquivos, bibliotecas e museus – padrões de metadados e de catalogação como MARC (MARC STANDARDS, 2017), MODS, Dublin Core, RDA (2010), ISAD(G), LIDO, e modelos conceituais como FRBR (IFLA, 1998), CIDOC CRM (2014), Ric-CM (INTERNATIONAL COUNCIL ON ARCHIVES, 2016), EDM (EUROPEANA DATA MODEL DOCUMENTATION), possuem diferentes exemplos de *relações* entre objetos de acervos. O papel das relações para contextualizar e tornar mais rica a navegação pelos registros de catálogos é ressaltada por Tillet (2001). Essas relações já existentes nos catálogos poderiam, uma vez que estes acervos fossem publicados segundo as tecnologias LOD, gerarem “links” entre objetos de acervos distribuídos de arquivos, bibliotecas e museus e também “links” para e de outros recursos na Web, como universidades, enciclopédias, agências de turismo, eventos, etc., enriquecendo mutuamente esses acervos e integrando-os aos fluxos de informação da Web.

Alguns exemplos da exploração das relações existentes nos padrões de descrição das três áreas que poderiam gerar “links” são: relações como um objeto e seu autor (campo 100 do MARC), entre uma publicação e seu ilustrador (campo 245 do MARC), entre uma obra de arte e suas versões segundo outros artistas (relação FRBR *is\_an\_adaptation\_of*).

## 5. Considerações Finais ou Conclusões

As tecnologias LOD colocam a questão da publicação e interoperabilidade de acervos de instituições de memória e cultura em um novo patamar (ACENJO-BULLÓN, 2015). Uma vez publicados, podem ser feitos “links” interligando diferentes conteúdos e assim agregando valor aos mesmos. Arquivos, bibliotecas e museus muito frequentemente possuem acervos complementares. Essas instituições podem usar as tecnologias LOD para tirarem partido dessas sinergias e complementariedades.

## 6. Referências:

- ACENJO-BULLÓN, X. Bibliotecas virtuales en 2014, año de consolidación del modelo *linked open data*. Anuário Think EPI, p. 9-21, 2015. Disponível em: <<http://dx.doi.org/10.3145/infor.2015.02>>. Acesso em 11 abr. 2017.
- BERNERS-LEE, T. *Linked data*. 2006. Disponível em: <<https://www.w3.org/DesignIssues/LinkedData.html>>. Acesso em: 20 fev. 2017.
- THE CIDOC CONCEPTUAL REFERENCE MODEL, version 5.1.12, January 2014. Disponível em: <[http://www.cidoc-crm.org/docs/cidoc\\_crm\\_version\\_5.1.2.pdf](http://www.cidoc-crm.org/docs/cidoc_crm_version_5.1.2.pdf)>. Acesso em 15 nov. 2014.
- DEMPSEY, L. The Library Catalogue in the New Discovery Environment: Some Thoughts. *Ariadne*, v. 48, July 2006. Disponível em <<http://www.ariadne.ac.uk/issue48/dempsey/>>. Acesso em: 1 julho 2009.
- EUROPEANA DATA MODEL DOCUMENTATION. [s.d.] Disponível em: <<http://pro.europeana.eu/page/edm-documentation>>. Acesso em 15 nov. 2014.
- GORMAN M. Cataloguing in an electronic age. *Cataloging & classification quarterly*, v. 36, n. 3-4, 2003. Disponível em: <<http://www.columbia.edu/~rjb57/electronic.pdf>>. Acesso em: 1 julho 2009.
- HANNEMANN, Jan. *Linked Data for Libraries*. In: WORLD LIBRARY AND INFORMATION CONGRESS: 76TH IFLA GENERAL CONFERENCE AND ASSEMBLY, 10-15 August 2010, Gothenburg, Sweden, Proceedings... Disponível em: <<http://www.ifla.org/files/hq/papers/ifla76/149-hannemann-en.pdf>>. Acesso em: 4 fev. 2012.
- HILLMANN, Diane I. The Emerging Cataloging Future. *Technicalities*, v. 28, no. 4, July/Aug. 2008. Disponível em <<http://hdl.handle.net/1813/10835>>. Acesso em 1 julho 2009.
- HANNEMANN, J. *Linked Data for Libraries*. In: WORLD LIBRARY AND INFORMATION CONGRESS: 76TH IFLA GENERAL CONFERENCE AND ASSEMBLY, 10-15 August 2010, Gothenburg, Sweden, Proceedings... Disponível em: <<http://www.ifla.org/files/hq/papers/ifla76/149-hannemann-en.pdf>>. Acesso em: 4 fev. 2012.
- HEATH, T.; BIZER, C. *Linked Data: Evolving the Web into a Global Data Space* (1st edition). *Synthesis Lectures on the Semantic Web: Theory and Technology*, v.1, n.1, p.1-136, 2011. Disponível em: <<http://linkeddatatobook.com/editions/1.0/>>. Acesso em: 15 ago. 2016.
- HILLMANN, D. I. The emerging cataloging future. *Technicalities*, July/Aug. 2008. Disponível em: <[https://www.researchgate.net/profile/Diane\\_Hillmann/publication/277216971\\_The\\_Em](https://www.researchgate.net/profile/Diane_Hillmann/publication/277216971_The_Em)

erging\_Cataloging\_Future/links/562f7a7308ae4742240ae304.pdf>. Acesso em: 14 jun. 2013.

IFLA Study Group on Functional Requirements for Bibliographic Records. Functional Requirements for Bibliographic Records: final report. München: K. G. Saur, 1998. INTERNATIONAL COUNCIL ON ARCHIVES. Experts group on archival description. Records in context a conceptual model for archival description. (Consultation Draft v0.1). ICA, 2016. Disponível em: <<http://www.ica.org/sites/default/files/RiC-CM-0.1.pdf>>. Acesso em: 23 nov. 2016.

IRI - Internationalized Resource Identifier. Disponível em: <<https://pt.wikipedia.org/wiki/IRI>>. Acesso em: 13 mar. 2017.

LIBRARY LINKED DATA INCUBATOR GROUP FINAL REPORT. W3C, 2011. Disponível em: <<http://www.w3.org/2005/Incubator/ld/XGR-ld-20111025/>>. Acesso em: 3 fev. 2012.

LIDO, What is LIDO, Lightweight Information Describing Objects. [s.d.]. Disponível em: <<http://network.icom.museum/cidoc/working-groups/lido/what-is-lido/>>. Acesso em: 8 jun. 2017.

MARC STANDARDS, Library of Congress, 2017. Disponível em: <<https://www.loc.gov/marc/>>. Acesso em: 23 jun. 2017.

MANN, T. The Changing Nature of the Catalog and Its Integration with Other Discovery Tools. Final Report: March 17, 2006. Prepared for the Library of Congress by Karen Calhoun. Disponível em: <<http://dx.doi.org/10.1080/10911360802087374>>. Acesso em: 13 mai. 2015.

MARCONDES, C. H. Interoperabilidade entre acervos digitais de arquivos, bibliotecas e museus: potencialidades das tecnologias de dados abertos interligados. *Perspectivas em Ciência da Informação*, v.21, n. 2, 2016. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/2735>>. Acesso em: 2 dez. 2016.

MODS Metadata Object Description Schema. Library of Congress, 2017. Disponível em: <<http://www.loc.gov/standards/mods/>>. Acesso em: 23 jun. 2017.

RDA: Resource Description and Access. 2010. Disponível em: <<http://www.rda-jsc.org/rda.html>>. Acesso em: 8 abr. 2012.

RDF PRIMER. MANOLA, F.; MILLER, E.(eds.). W3C, 2004. Disponível em: <<http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>>. Acesso em: 15 dez. 2011.

TILLET, B. B. Bibliographic relationships. In: Bean, C. A. and Green, R. (Eds.), *Relationships in the Organization of Knowledge*. Dordrecht: Kluwer Academic Publishers, 2001. p. 19-35.

W3C Library Linked Data Incubator Group. Datasets, Value Vocabularies, and Metadata Element Sets, 2011. Disponível em: <[http://www.w3.org/2005/Incubator/ld/wiki/Vocabulary\\_and\\_Dataset](http://www.w3.org/2005/Incubator/ld/wiki/Vocabulary_and_Dataset)>. Acesso em: 31 de março de 2012.

### **Agências financiadoras**

Projeto apoiado pelo CNPq através de bolsa de produtividade